

The copular subschema [*become/devenir* + past participle] in English and French: productivity and degrees of passivity¹

Niek Van Wettere

Vrije Universiteit Brussel & Universiteit Gent

This article presents a contrastive analysis of the English copular subschema [*become* + past participle] and the equivalent copular subschema [*devenir* + past participle] in French, based on web data. It is shown that both patterns are almost equally productive at the subject complement level. Furthermore, a more in-depth analysis demonstrates that, in the segment of participles with a high adjectival potential, *devenir* accumulates more participle tokens than *become*. Conversely, the reverse holds true for participles with a high verbal potential, in which case *become* is characterized by more participle tokens than *devenir*. This high amount of combinations between *become* and eventive participles also suggests a higher degree of passivity for *become*. However, in the segment of participles with an intermediate verbal potential, *devenir* is slightly more type frequent than *become*, which hints at an emerging productivity in this area for *devenir* as well.

Keywords: *become*, copular construction, passive construction, productivity, English/French

1. Introduction

This article presents a contrastive analysis of two copular verbs, namely English *become* and French *devenir*, which are the prototypical copulas expressing change-of-state in their respective languages. More specifically, this article focuses on the subschemas [*become* / *devenir* + past participle], for example *become imprisoned* or *devenir interdit* ('become prohibited'). The notion of 'subschemas' is to be interpreted within the paradigm of Construction Grammar (Goldberg, 1995), which postulates a taxonomic continuum from substantive

¹ I would like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper.

constructions (i.e. all the slots are lexically filled) to completely schematic constructions. Subschemas are more specific than the overarching completely schematic construction, but less so than the substantive constructions. In the present case, the subschemas [*become* / *devenir* + past participle] are encompassed by the more general copular constructions [*become* / *devenir* + subject complement].

English and French behave similarly with regard to this subschema, in the sense that the subschemas [*become* / *devenir* + past participle] are in principle copular and not sanctioned by the passive construction. In contrast, the same pattern does automatically trigger a passive reading in German (cf. *werden*) and Dutch (cf. *worden*) (see example 1).

- (1) Goethes Name wird mit "oe" geschrieben. ('Goethe's name is written with "oe"')

Consequently, the current study can contribute significantly to our understanding of how this pattern is functionally implemented in different languages. Presumably, this pattern can be positioned cross-linguistically and diachronically on a continuum from copula construction to passive construction.

The objectives of our analysis are two-fold:

- to identify possible productivity differences between the subschemas [*become* + past participle] and [*devenir* + past participle], despite the fact that they are arguably the closest functional equivalents in the context of a comparison between English and French, given that they both fulfil the role of prototypical change-of-state copula in their respective languages;
- to examine which verb, *become* or *devenir*, has the highest degree of passivity, i.e. a global distributional profile that displays more passive-like characteristics. These passive-like characteristics will be mainly assessed through the adjectival potential of the participles combining with *become* or *devenir*. Since the same pattern functions as a passive construction in languages such as German and Dutch, this is a relevant research question to put forward, even if none of our corpus examples can actually be recognized as a full-fledged passive.

With regard to the first objective, it can be hypothesized that *become* has a higher productivity than *devenir* in the area of participial subject complements. In the literature, it has been observed that *devenir* is to a large extent incompatible with past participles (Guehria, 2011:139). According to Guehria (2011), this incompatibility is caused by the non-gradability that characterizes these participles in the position of subject complement. Conversely, if the a priori non-gradable participle is coerced into a gradable use under the influence of, for example, a scalar construction (cf. [*de moins en moins* X], 'less and less X'), this

improves the acceptability of the combination with *devenir*, as shown in example (2).

- (2) Hélas, avec les récents propos du Président, il est à craindre que ce texte classique *deviendra de moins en moins analysé* dans les classes. (Guehria, 2011:143 ; ‘become less and less analyzed’)

However, this line of reasoning seems flawed, because *devenir* can very well combine with non-gradable adjectives such as *obligatoire* (‘obligatory’) and *responsable* (‘responsible’). Instead, it might be the case that *devenir* is simply reluctant to combine with ‘more verbal’ participles, whereas the combination with ‘more adjectival’ participles is more straightforward.² In the same vein, these more adjectival participles are often gradable in nature, which explains the analysis developed in Guehria (2011). The combination [*devenir* + more adjectival, gradable participle] is illustrated in example (3).

- (3) C'est fou comment tu pouvais *devenir intéressé à ce cours* tout d'un coup. (Google, men's magazine, ‘become interested’)

In contrast, it can be assumed that *devenir* is less compatible with more verbal participles, even though there is some evidence to the contrary (see examples 4 and 5).

- (4) Au cours de la nuit, le bâtiment se congèle et *devient emprisonné* dans un cube de glace. (Google, art museum brochure, ‘become imprisoned’)
- (5) Par contre, à l’instar de tous les matériaux exposés aux intempéries, l’aluminium peut *devenir endommagé* et finir par se détériorer. (Google, renovation guidelines, ‘become damaged’)

Contrary to *devenir*, *become* seems more predisposed to combine with more verbal participles, derived from action verbs such as *catch* (example 6) and *bury* (example 7).

- (6) [...] when the mammals *became caught* in the fishing nets. (BNC)
- (7) A few days later, the storm became so violent that sheep *became buried* in six-foot drifts of snow, [...]. (BNC)

The evidence presented here is of course anecdotal: this will be examined more thoroughly in this paper.

In sum, the alleged incompatibility of *devenir* with (more verbal) past participles, as described by Guehria (2011), as well as the apparent ease with which *become* allows for such a construction, warrant a null hypothesis that stipulates a higher productivity for *become* than for *devenir*. If *become* is indeed

² The notions of adjectival and verbal participle will be more concretely operationalized in Section 4.

more receptive towards more verbal past participles than *devenir*, this would also hint at a higher degree of passivity for *become*. The passive construction is highly applicable to prototypical action verbs, and less so to, for example, psych verbs.³ Logically, the higher the verbal potential of the past participle in question and, conversely, the lower its adjectival potential, the more likely it becomes that [*become* + past participle] can be construed as a passive. Since other Germanic languages such as German (cf. *werden*) and Dutch (cf. *worden*) do use the prototypical copula that denotes change-of-state as auxiliary of the passive voice, English *become* can be expected to have at least some passive-like features.

The contrastive analysis between *become* and *devenir* is implemented by conducting an extensive web corpus study, adopting a usage-based framework. The remainder of the article is structured as follows. First, the general methodology is outlined in Section 2, which presents the composition of the sample. In Section 3, the productivity of *become* and *devenir* will be assessed at the level of the participial subject complement. Next, the degree of passivity will be examined for both verbs (cf. Section 4), through the prism of the adjectival potential of the past participles with which *become* and *devenir* combine. The main conclusions that can be drawn from this study will be summarized in Section 5.

2. Composition of the web data sample

In order to make the contrastive comparison between *become* and *devenir*, two samples, consisting each of 10,000 occurrences, were randomly extracted from the English Web Corpus enTenTen 2013 and the French Web Corpus frTenTen 2012, respectively. Both corpora belong to the TenTen Corpus Family (Kilgarriff *et al.*, 2014), which guarantees a relatively solid basis for cross-linguistic comparison.⁴ The following search queries were used to identify the relevant pattern [*become* / *devenir* + past participle]: “[lemma = “become”] [tag = “VVN”]” and [lemma = “devenir”] [tag=“VER:pper”].

Subsequently, the relevant copula uses within each sample were identified and selected up to 2,500 instantiations, which results in a total sample size of 5,000 corpus examples for both verbs. The following cases were discarded:

- Nonsensical examples, containing random words.

³ The distinction between action and non-action (“kinesis”) is one of the parameters that determines the degree of transitivity of a verb (Hopper and Thompson, 1980). High transitivity (~ action verb) enables passivation.

⁴ Since all the corpora from the TenTen family are crawled from the internet by means of the same Spiderling tool, it can be expected that the corpora are comparable. Of course, it cannot be excluded that the English and French internet are differently structured in terms of language varieties, distribution of genres etc., but it seems unlikely that these differences would be substantial enough to significantly distort the picture of the constructions examined here.

- Translations, especially machine-translated data, to the extent that these could be identified as such by the researcher.
- Participles without a clear link with a corresponding infinitive, either because the meaning of the corresponding infinitive is too different from the meaning of the participle (in the construction [*become/devenir* + past participle]) (e.g. *cinglé* ‘crazy’, vs. the corresponding verb *cingler* ‘hit’) or because there is no corresponding infinitive, at least in synchrony (e.g. *enceinte* ‘pregnant’).⁵
- French participles that have an adjectival as well as a nominal interpretation with respect to an animate subject, e.g. (*un*) *associé*. Since English participles do in general not have the same twofold interpretation (cf. ‘associated’ vs ‘an associate’), these French cases were excluded from the study, in order to ensure a valid comparison between the two languages.
- English participles that contain a particle, for example *fleshed out*. Since particles are generally not present in French, these English cases were excluded from the study.

Importantly, the verbs from which the included participles are derived can be of any prototypical valency (monotransitive, ditransitive, etc.). In principle, only participles of which the corresponding verbs prototypically adopt the schema [subject + verb + direct object] in the active voice are eligible to receive a passive interpretation in combination with the passive auxiliary *be/être*. On the basis of this final sample of 5,000 occurrences, an analysis of the productivity and the degree of passivity is presented in Sections 3 and 4.

3. General productivity analysis

In the first stage of our analysis, the productivity of *become* and *devenir*, measured at the level of the subject complement slot, is contrasted. In order to assess the productivity of both constructions, a series of well-known productivity measures is used, as defined in the work of Baayen (2009) and applied to syntactic constructions by Zeldes (2012). Since the sample size is in both cases equal, the obtained results are comparable (Gaeta and Ricca, 2006).

Table 1 gives an overview of the different measurements, namely type frequency, hapax⁶ frequency, dis legomena⁷ frequency, hapax/type ratio, dis

⁵ This also encompasses certain cases where the participle is preceded by a prefix, e.g. ‘immuno-’ in *immunodéprimé*, in which case there is no corresponding infinitive ‘immunodéprimer’.

⁶ Hapax legomena correspond to types occurring only once in the sample.

⁷ Dis legomena are types that occur twice in the sample.

legomena / type ratio and highest token frequency among the different types. These measurements were computed for both samples of 2,500 occurrences. Note that a drawback of this approach compared to an approach that calculates these measures for the complete non-sampled source corpus is that the difference in occurrence rate of the subschemas is not factored in (Baayen, 1993:206). As it turns out, such a difference might very well apply. The Sketch Engine search query indicates that the subschema [*become* + past participle] (ca. 38 occurrences per million words) seems more frequent than its French counterpart (ca. 9 occurrences per million words), which suggests a higher degree of passivity for the English pattern. Since these overall results are not manually cleaned, they are of course only approximate.

Table 1. Productivity comparison of *become* and *devenir*.

verb	sample size	type frequency⁸	hapax frequency	dis legomena frequency
become	2500	638	351	104
devenir	2500	674	352	125
verb	hapax / type ratio	dis legomena / type ratio	highest token frequency among the different types	
become	0.55	0.16	288	
devenir	0.52	0.19	269	

The results of *become* and *devenir* shown in Table 1 are strikingly similar for all the productivity measures included. On this point, our initial hypothesis that *become* is characterized by a higher degree of passivity than *devenir* does not seem to be borne out. Indeed, if this hypothesis were true, *become* would be more productive than *devenir* in the area of participial subject complements, which is not the case.

Next, the Type-Frequency List plot (Baroni and Evert, 2014) in Figure 1 zooms in on the 50 types with the highest token frequencies, ranked from high to low. This enables a better view on the shape of the ‘frequency summit’, i.e. the most frequent types. This frequency summit sheds light on the extent to which the subschemas contain one or multiple conventionalized types. Conventionalization can be viewed as detrimental to productivity. Again, it can be observed that *become* and *devenir* behave in a very similar manner.

⁸ It is important to note that possible polysemy of the participles was not accounted for in the process of determining the different types.

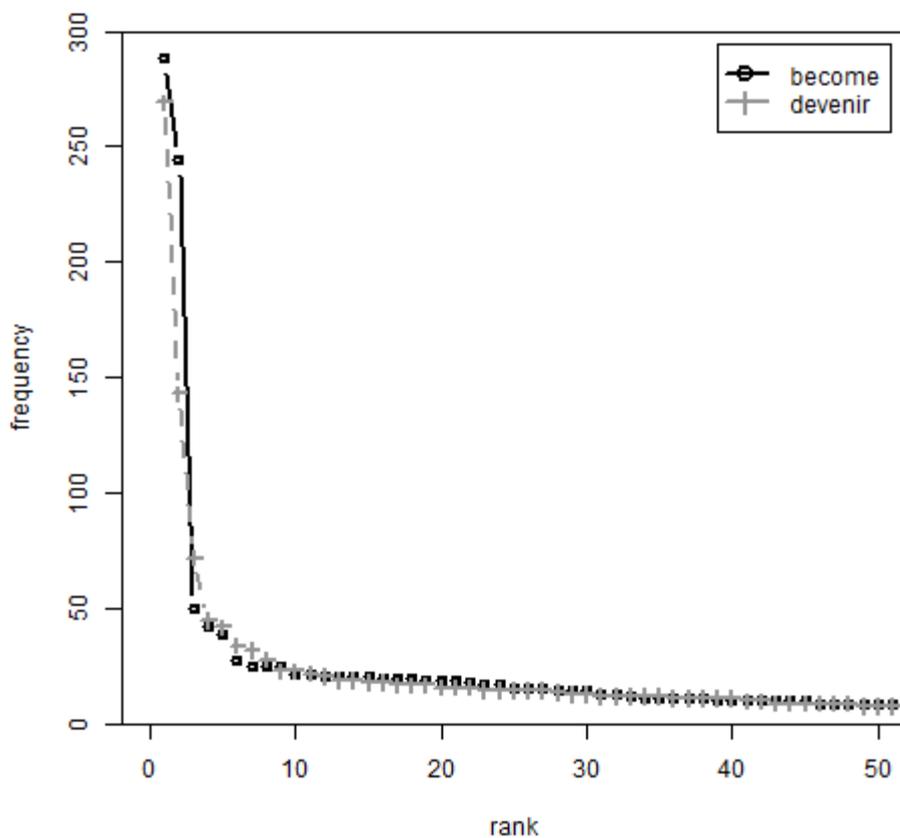


Figure 1. Type-Frequency List.

The only difference is that *become* is characterized by two extremely frequent types (*known*: 288 tokens ; *involved*: 244 tokens), whereas the difference between the first rank and the second rank is more important in the frequency distribution of *devenir* (*compliqué*: 269 tokens ; *connu*: 143 tokens).

The aforementioned ‘static’ observations can be complemented by a more dynamic view, which tracks the evolution of a certain statistic throughout the sample. The plot in Figure 2 shows the empirical vocabulary growth (Baroni and Evert, 2014), namely the evolution of the type frequency (= $V(N)$) within the window [1 ; 2500 corpus examples] for *become* and *devenir*. The two lowest lines, which are thinner than the two lines in the upper part of the plot, represent the evolution of the hapax frequency (= $V_1(N)$).

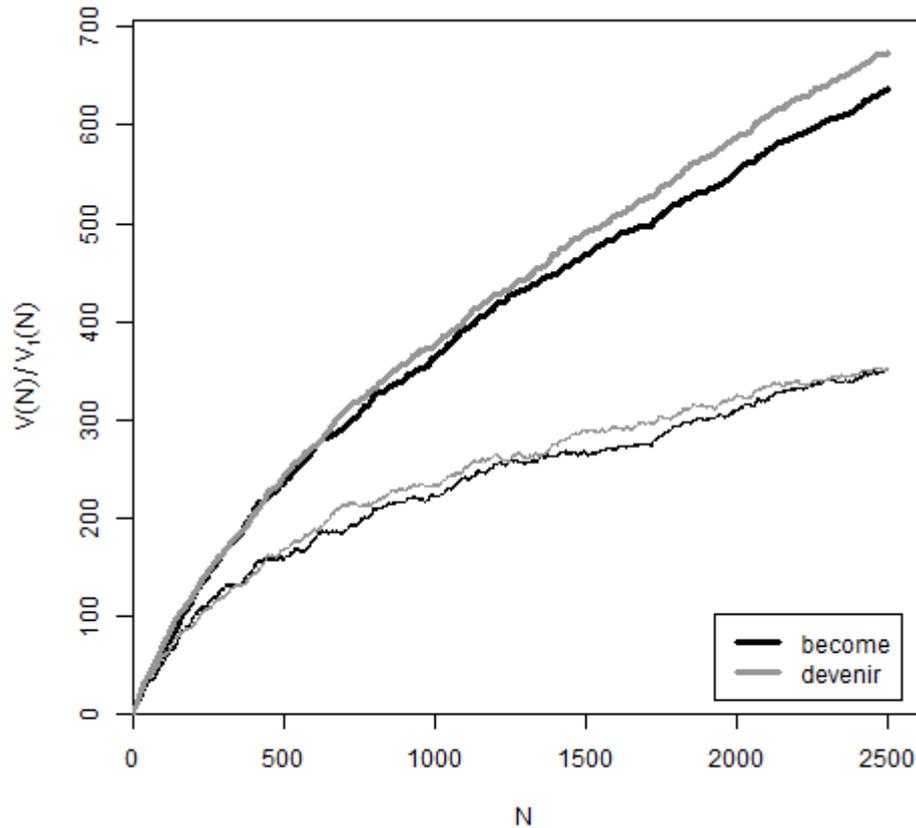


Figure 2. Empirical Vocabulary Growth

Since a statistic in isolation is rather uninformative about the anticipated evolution of the statistic beyond the attested sample size, it is important to examine the global tendency of the statistic. Once more, the empirical vocabulary growth plot confirms that *become* and *devenir* are very much alike: the curves almost coincide, especially on the level of the hapax frequency, a measure which captures the most important aspect of productivity, namely potential productivity. At the population level, potential productivity denotes the likelihood with which a schema will extend its scope to new types. As Zeldes (2012) highlights, “it is the coining of new forms that constitutes productive usage” (Zeldes, 2012:37). At the sample level, potential productivity also estimates the degree to which the sample of a given size exhausts all the types present in the population (Baayen and Lieber, 1991:837). Contrary to the hapax frequency, the type frequency curves display a slightly widening gap in favour of *devenir*, which could widen even more beyond the attested sample size.

In conclusion, the productivity analysis has shown that *become* and *devenir* are almost identical in this respect. Of course, the approach adopted in this section is holistic. In the next section, certain subsets of participles are detected according to two quantitative parameters. In this way, it is possible to zoom in on certain areas within the subject complement slot that are especially relevant to the

comparison at hand, most notably concerning the degree of passivity of *become* and *devenir*.

4. Degree of passivity measured by means of the adjectival potential of past participles

Up to this point, the analysis did not expose any major differences between the subschemas [*become* + past participle] and [*devenir* + past participle]. From a general productivity perspective, *become* and *devenir* seem to operate at (more or less) the same level of openness, determined by means of the type frequency, and extensibility, as measured by the hapax frequency. According to the initial research hypothesis formulated in Section 1, the subschema [*become* + past participle] is more likely to contain instantiations that tend towards the passive construction than [*devenir* + past participle]. If this prediction does not bear out in terms of a higher overall productivity for [*become* + past participle] in the samples analyzed, other indicators pertaining to the opposition between verbal and adjectival participles may corroborate this hypothesis.

This section aims to examine the ‘adjectival potential’ of the past participles that combine with *become* and *devenir*. The notion of ‘adjectival potential’ is addressed in De Sutter (2005:225) and denotes the potential of the past participle to function as an adjective. This potential is determined out of context, based on a set of generic distributional traits. More specifically, two distributional traits⁹ are studied more closely in the next two subsections for the participles combining with *become* and *devenir*, namely:

- i. Possibility of modification of the past participle by a degree adverb (cf. Section 4.1). If the participle is frequently modified by a degree adverb in usage (e.g. *very* in English and *très* in French), this is to be taken as evidence for the predominantly adjectival nature of the participle. Of course, modification by a degree adverb can only apply to those cases that are intrinsically gradable. The non-gradable participles that are nonetheless adjectival do not fall under the scope of this test. In other words, participles for which the ratio¹⁰ [token frequency of the participle accompanied by degree adverb modification / token frequency of the participle, all

⁹ Note that there are additional distributional traits that contribute to the adjectival potential of participles, such as the possibility of prefixation by [*in-* / *un-*], for example: *unmodified* (< *modified*).

¹⁰ Given that only seven participles have a token frequency in the TenTen corpora that is less than 100 and that the vast majority have a token frequency higher than 1000, the ratios should be reliable.

contexts] is higher should be more inclined towards an adjectival use.

- ii. Frequency of the past participle form in comparison with other verbal forms (cf. Section 4.2). If, for a given verb, the overall frequency of the corresponding past participle is relatively high compared to the overall frequency of other forms of the verb that are unequivocally verbal (e.g. infinitive), this implies that the past participle is conventionalized and, consequently, has to a certain extent emancipated itself from the rest of the verbal paradigm. It is often the case that an established adjectival use of the participle in question is responsible for this conventionalization. In addition to their verbal use, these participles are able to occur frequently in prototypically adjectival positions (subject complement, noun modifier, etc.), which are less open to predominantly verbal participles. In sum, it can be hypothesized that participles with a proclivity for adjectival use will have a low ratio [frequency of the infinitive / frequency of the participle¹¹] compared to participles that are mainly restricted to a verbal use.

It is noteworthy that a Spearman correlation analysis between the variables that operationalize (i) and (ii) above, further explained in Sections 4.1 and 4.2, highlights that both variables are negatively correlated. This is unsurprising, since a high value for ratio (i) points towards a strong adjectival use, whereas a high value for ratio (ii) indicates a predominantly verbal use for the participle in question. Adjectival and verbal use being each other's opposites, a negative correlation makes sense. However, the negative correlation is fairly modest (-0.24; p -value = 0), which suggests that both dimensions of variation are not equivalent. Consequently, it is not redundant to include both indicators in our analysis.

How does this tie in with the degree of passivity of the two subschemas [*become* + past participle] and [*devenir* + past participle]? It is known that the adjectival potential of participles can contribute to the disambiguation between the copular construction involving *be* as copula verb and the auxiliary construction *be*, comprising both the passive construction and the compound tense. The former supposes an adjectival use of the participle, whereas the latter entails a verbal use of the participle. At the surface, the schema [*be* + past participle] is identical in both cases, but the functional analysis of the participle is different: subject complement vs complement of the auxiliary. This can be illustrated by means of examples 8 and 9.

¹¹ The frequency of the participle is placed in the denominator because this value cannot be zero in this study (otherwise, the participle would simply not be part of our dataset).

- (8) She is motivated.
(9) The door is opened.

Assuming that the participle *motivated* has, out of context, a higher adjectival potential than the participle *opened*, it can be deduced that example (8) is more likely to be parsed as [copula verb *be* + adjectival subject complement *motivated*] and example (9) is more likely to be interpreted as [auxiliary *be* + verbal complement *opened*]. In other words, the adjectival potential of the participle can determine the constructional parsing.

This line of reasoning can now be extended to the subschemas [*become* + past participle] and [*devenir* + past participle]. Contrary to the pattern [*be* + past participle], there is in principle no a priori ambiguity concerning the functional analysis of these two patterns: both *become* and *devenir* are in traditional grammars only recognized as copula verb and not as passive auxiliary. Nevertheless, it is noteworthy that, from a typological perspective, German *werden* and Dutch *worden*, the closest equivalents of *become* and *devenir*, do take on the role of passive auxiliary, in addition to their use as prototypical change-of-state copula verb. This association between copula and auxiliary in synchrony is also reflected by the well-known diachronic pathway from copula to auxiliary (Dik, 1987; Laca, 2000; Sansò and Ramat, 2015). Consequently, the extent to which copulas *become* and *devenir* display certain passive characteristics in their combination with past participles merits investigation and can possibly highlight ongoing passive auxiliarization in synchrony. Since the existence of polygrammaticalized items (Craig, 1991) that function both as change-of-state copula and as passive auxiliary seems mainly attested in Germanic languages (e.g. German and Dutch) and not in Romance languages (e.g. Spanish and Italian), it follows that English *become* is probably more likely to possess passive-like characteristics than French *devenir*.

One of the ways to assess the degree of passivity is through the adjectival potential of the past participles that combine with *become* and *devenir*. A low adjectival potential increases the likelihood of a verbal use of the participle. It goes without saying that a participle is used verbally in the passive construction. However, if a participle has a low adjectival potential, this does not automatically entail that the subschema [*become* + past participle] is not an instantiation of the copula construction. Still, a higher general preference for verbally used participles, even within the context of a copula construction, may be a precursory sign of a future development towards passive auxiliary.

Finally, it is important to note that the parallel between [*be* + past participle] and [change-of-state copula + participle] is not flawless. German *werden* and Dutch *worden* unambiguously instantiate in the vast majority of cases the passive construction when combined with a past participle. It follows from this that the same constructional ambivalence that affects [*be* + past participle] does not apply

to *werden* and *worden*. However, this does not mean that the pattern [change-of-state copula + participle] can never exhibit this ambivalence between copula construction and auxiliary construction. For example, as shown in Table 2, the English verb *get* can occur both in the copular and passive constructions.

Table 2. Copular and passive constructions of *get* (English).

passive construction	
[<i>get</i> + past participle]	It happened in the wing's kitchen when a prison officer suddenly got attacked by an inmate. (Google)
copular construction	
[<i>get</i> + participial subject complement]	She was looking for Chris and got depressed after a while. (Google)
[<i>get</i> + adjectival subject complement]	I got pregnant , then a Wall Street firm pushed me out of my job. (Google)

In sum, not only can a prototypical change-of-state copula also occur in the passive construction (cf. German *werden* and Dutch *worden*), change-of-state semi-copulas such as *get* demonstrate that the pattern [change-of-state copula + past participle] can have a constructional ambivalence similar to the one attested for [*be* + past participle]. This adduces further evidence to the necessity of an in-depth analysis of the subschemas [*become* + past participle] and [*devenir* + past participle]. The case of *get* is also important because of the fact that French seems to lack a direct equivalent of *get*, presenting the same constructional ambivalence. Again, this lends further support to our hypothesis that English *become* is more likely to have passive-like characteristics than French *devenir*.

Sections 4.1 and 4.2 explore the two aforementioned distributional traits assessing the adjectival potential of the participle. *Become* and *devenir* are contrasted with each other in order to determine which of the two verbs has an overall profile that is most oriented towards verbal participles. This proclivity for verbal participles will be taken as evidence for a higher degree of passivity, for the reasons explained above. It must be stressed that other factors than the two distributional traits addressed here are susceptible to influence passivity in context (for a comprehensive overview, cf. De Sutter, 2005; Raineri, 2010), but a full quantitative account of all these elements is out of the scope of this study.

4.1 Possibility of modification of the past participle by a degree adverb

First, the possibility of modification of the past participle by a degree adverb is examined. This is operationalized as the measure ‘token frequency of [*very/très*

+ participle] / token frequency of [participle]¹², multiplied by 100, for which the token frequencies are calculated based on the complete source corpora, namely the English Web Corpus (enTenTen 2013) and the French Web Corpus (frTenTen 2012). Since this measure corresponds to a ratio, the results of *become* and *devenir* can be compared, in spite of the different size of the source corpora (19 billion words for English and 10 billion words for French). A high ratio can be interpreted as a sign of high adjectival potential.

Table 3 lists the 20 participles which obtained the highest ratios. Since this table contains the types with the highest adjectival potential in our data, it is to be expected that certain types are very close to full-fledged adjectives. In the same vein, the relationship with the corresponding infinitive can be rather distant in cases where the infinitive is much less cognitively entrenched (e.g. *doué* vs. *douer*).

Table 3. Top 20 of highest ratios ‘token frequency of [*very/très* + participle] / token frequency of [participle]’.

rank	verb	participle	ratio
1	devenir	prisé (‘popular’)	40.5
2	devenir	controversé (‘controversial’)	16.8
3	devenir	convoité (‘coveted’)	15.1
4	devenir	doué (‘gifted’)	13.9
5	devenir	répandu (‘widespread’)	12.9
6	devenir	typé (‘typed’)	11.8
7	become	impressed	11.6
8	devenir	varié (varied)	10.6
9	devenir	diversifié (‘diversified’)	10.2
10	become	excited	9.9
11	devenir	apprécié (‘appreciated’)	9.6
12	devenir	serré (‘tight’)	8.9
13	devenir	politisé (‘politicised’)	8.4
14	devenir	impressionné (‘impressed’)	8.3
15	devenir	vallonné (‘hilly’)	8.2
16	devenir	éloigné (‘far’)	8.0
17	devenir	affûté (‘honed’)	7.7
18	devenir	compliqué (‘complicated’)	7.4
19	devenir	excité (‘excited’)	7.1
20	devenir	coloré (‘colored’)	7.1

Clearly, the majority of these highest ranked participles combine with *devenir*. This is a first confirmation of our initial hypothesis that the participial subject

¹² Note that, for French, the token frequencies were only calculated for the masculine form of the participle. Since this applies to both the numerator and the denominator, the ratio should still be a reliable estimate.

complements combining with *devenir* generally have a higher proclivity for predominantly adjectival participles than those combining with *become*. Of course, it has to be borne in mind that the measure used here assesses adjectival potential only indirectly, through an assessment of the modifiability and intrinsic gradability of the participle.

Next, the distribution of this measure in our samples of *become* and *devenir* is evaluated by means of a frequency polygon graph (Wickham, 2009: 68-72). This type of plot essentially corresponds to a histogram, but the bins are left out, making it easier to compare two or more distributions. In addition to the frequency polygons, the 75th percentile of the values observed for *become* and *devenir* is also indicated on the plot by means of a vertical line. Importantly, the plot captures the entire dataset and is not solely based on the attested types and their corresponding values. In other words, the token frequency of the participle within the sample is also taken into consideration: participles will have more impact on the distribution according to the frequency with which they occur in the dataset. Furthermore, the scope of the x-axis, which indicates the aforementioned ratio, has been restricted to a reduced range (from ca. 41 to ca. 10), so that 22 outliers are not visualized within the limits of the plot. The comparison of *become* and *devenir* is visualized in Figure 3.

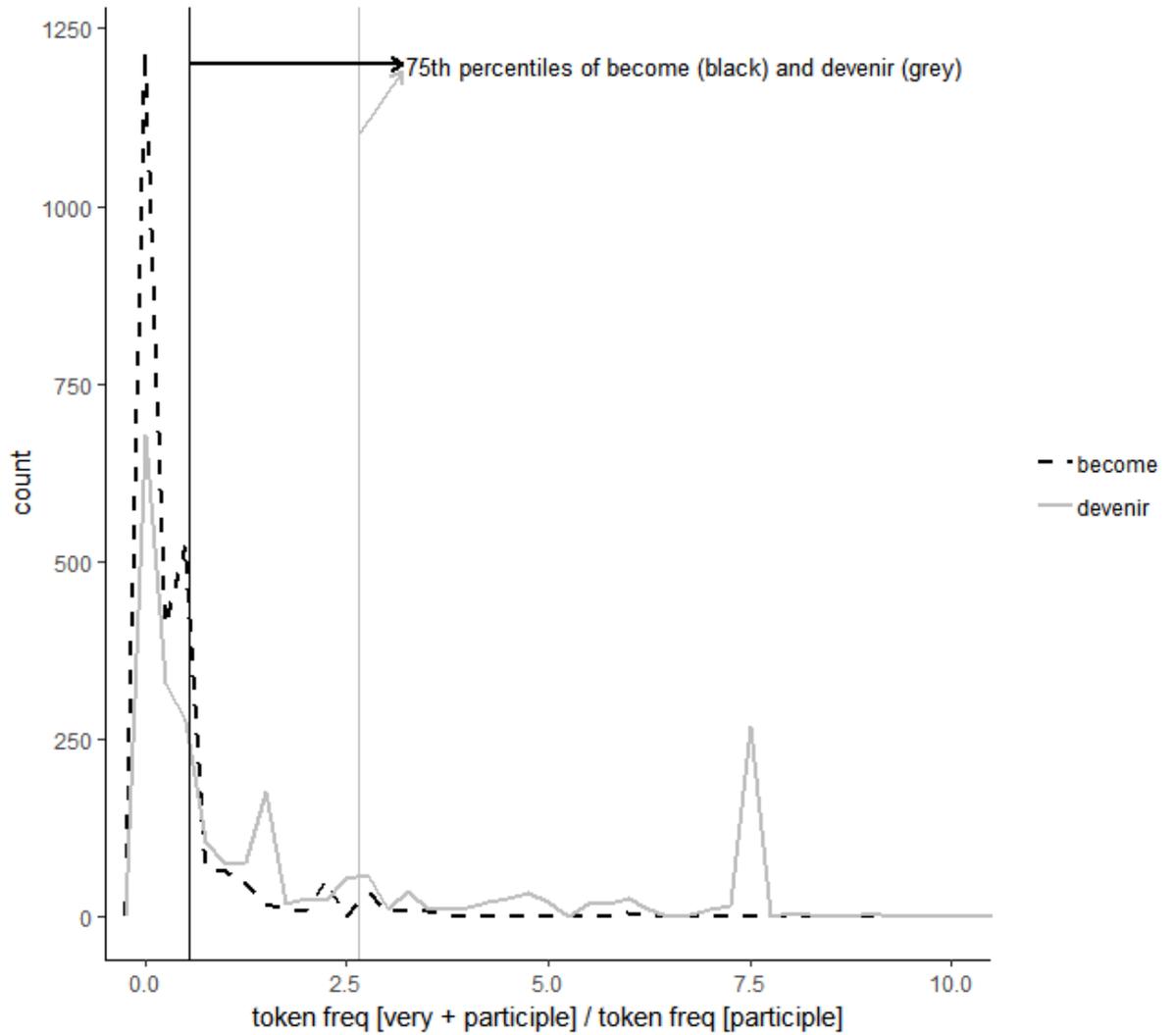


Figure 3. Frequency polygon graph with comparison of *become* and *devenir* for the ratio ‘token freq [*very/très* + participle] / token freq [participle]’.

This plot demonstrates that *devenir* accumulates more participle tokens than *become* in the upper segment of ratio values. This is evidenced by the fact that *become* reaches a higher peak in the lower segment of ratio values between 0 and 0.5 than *devenir*. Inversely, the plot indicates very distinctly two peaks in the distribution of *devenir* for ratio values higher than 0.5. This information can be complemented with a series of key summary statistics characterizing both distributions (Table 4).

Table 4. Summary statistics.

	mean	median	min	max	25 th perc.	75 th perc.
become	0.42	0.14	0	11.6	0.03	0.55
devenir	2.04	0.54	0	40.5	0.11	2.66

These data show that most of the ratio values attested for *become* remain fairly close to the median of 0.14, cf. the Inter Quartile Range (IQR) between 0.03 and 0.55. In contrast, the variability observed for *devenir* is much higher, with an IQR between 0.11 and 2.66.

Moreover, Table 5 reports the type frequency and the hapax frequency of the participles within three intervals of ratio values: [0 ; 75th percentile of *become* (0.55)],]75th percentile of *become* (0.55) ; 75th percentile of *devenir* (2.66)] and]75th percentile of *devenir* (2.66) ; ∞]. Note that the 75th percentiles are also indicated on Figure 3.

Table 5. Productivity measures per interval of ratio values and per verb.

verb	interval	token frequency (within sample)	type frequency	hapax frequency
become	[0 ; 0.55]	2103	522	301
devenir	[0 ; 0.55]	1252	427	239
become]0.55 ; 2.66]	320	100	45
devenir]0.55 ; 2.66]	634	162	81
become]2.66 ; ∞]	77	17	6
devenir]2.66 ; ∞]	614	86	33

This complementary view reveals that *devenir* also surpasses *become* in the higher intervals]0.55 ; 2.66] and]2.66 ; ∞] from the perspective of type and hapax frequency. To conclude, all evidence points in the same direction: based on the possibility of modification of the participle by a degree adverb, the profile of *devenir* is more oriented towards participles with high adjectival potential, compared to *become*.

Finally, it remains to be determined which types contribute the most to this peak observed for *devenir*. In other words, which very frequent types have relatively high ratio values? Relatively high ratio values can be defined, quite arbitrarily, as greater than the 75th percentile of *devenir*. Table 6 provides the top 20 of highest participle token frequencies, measured within our sample, for this subset of relatively high ratio values.

Table 6. Top 20 of highest participle token frequencies (within dataset), restricted to the subset of participles with high values for the ratio ‘token freq [very + participle] / token freq [participle]’.

rank	verb	past participle	ratio	token frequency (within sample)
1	devenir	compliqué (‘complicated’)	7.4	269
2	devenir	agité (‘agitated’)	4.8	22
3	become	frustrated	2.8	22
4	devenir	limité (‘limited’)	5.6	19
5	devenir	foncé (‘dark’)	3.2	19

6	devenir	fatigué ('tired')	4.4	18
7	devenir	risqué ('risky')	6.0	17
8	devenir	élevé ('high')	5.7	16
9	devenir	excité ('excited')	7.1	15
10	devenir	impliqué ('implicated')	4.0	12
11	devenir	déprimé ('depressed')	2.7	11
12	become	relaxed	3.3	10
13	become	intrigued	2.8	9
14	devenir	poli ('polite')	4.8	8
15	devenir	intéressé ('interested')	4.3	8
16	devenir	doué ('gifted')	13.9	7
17	become	upset	6.0	7
18	devenir	serré ('tight')	8.9	6
19	devenir	coloré ('colored')	7.1	6
20	devenir	attaché ('attached')	6.3	6

This analysis shows that the type *compliqué* (< *compliquer*) is a major contributor to the higher amount of tokens in the upper segment of ratio values, observed for *devenir*. By means of comparison, the English equivalent type *complicated* only represents 10 tokens in the sample of *become*.¹³ Unsurprisingly, the top 20 also contains numerous participles derived from psych verbs (*frustrated*, *agité*, *fatigué*, *excité*, *déprimé* etc.). It is known that participles derived from psych verbs are often stative (non-eventive) and are therefore less prone to induce a passive interpretation in case of a pattern that allows both for a copular and passive construction (e.g. [*be* + past participle]), compared to participles derived from prototypical action verbs (Sleeman, 2014). In addition, psychological states are prototypically gradable in nature, which makes them very compatible with degree adverbs. Hence, these participles score high on the parameter of adjectival potential.

4.2 Frequency of the past participle form in comparison with other verbal forms

In this second subsection, the predominance of the past participle form in comparison with other unequivocally verbal forms of the same type is addressed. This can be operationalized as the ratio measure ‘token frequency [infinitive] /

¹³ The ratio value for *complicated* is also relatively high (3.79), but not as high as for *compliqué*.

token frequency [participle]¹⁴, multiplied by 100, where the infinitive¹⁵ is considered to be a proxy for an undisputed verbal use. As was the case for the first ratio measure discussed in Section 4.1, the token frequencies are calculated based on the complete source corpora. Contrary to the first ratio measure, a high ratio can now be interpreted as a sign of low adjectival potential. Participles with a high ratio should be more likely to convey a passive-like meaning in combination with *become*, such as examples 10 and 11.

- (10) If possible, find an expert inside your area who's willing to *become interviewed* on your goods. (enTenTen, escort website)
- (11) [...] l'apparition de troubles des mois plus tard *devenant comptabilisés* comme "moins d'un jour après" l'inoculation du vaccin. (frTenTen, personal blog, 'become counted')

In Table 7, the 20 participles that obtained the highest ratios are enumerated.

Table 7. Top 20 of highest ratios 'token frequency [infinitive] / token frequency [participle]'.

rank	verb	participle	ratio
1	become	prevented	1106.0
2	become	accommodated	1007.5
3	become	conformed	1005.6
4	devenir	compté ('counted')	763.5
5	become	checked	711.7
6	become	understood	701.6
7	become	reconnected	637.5
8	become	bridged	633.4
9	become	skipped	602.4
10	devenir	culpabilisé ('made guilty')	533.4
11	become	stricken	523.9
12	become	distanced	522.5
13	become	legitimized	512.2
14	devenir	admiré ('admired')	471.6
15	devenir	su ('known')	460.6
16	become	unravelling	451.1
17	devenir	regardé	450.1
18	become	capitalised	392.6

¹⁴ Contrary to the ratio defined in the previous section, the value 'token frequency [participle]' is now, for the French participles, calculated based on all the forms of the participle (masculine singular and plural, feminine singular and plural). Otherwise, this ratio would underestimate in French the true weight of the participle relative to the infinitival form, compared to English, which only has one participle form.

¹⁵ Only the forms that are tagged as infinitive in the corpus (frTenTen: "VER:infi"; enTenTen: "VV") were taken into account. The infinitive corresponding to the participle might have different meanings, all of which are not necessarily equally relevant for the participle following *become/devenir*. Consequently, this introduces some imprecision in the measurement.

19	become	possessed	384.4
20	devenir	possédé ('possessed')	371.1

As anticipated, most of the participles with very high verbal potential listed in Table 7 combine with *become*. Two interesting cases among the English participles are *prevented* and *skipped* (see examples 12 and 13).

- (12) Programs which suggest you are able to shed plenty of bodyweight whilst consuming whatever meals you would like, rather than performing any kind of physical exercise, are to *become prevented*. (enTenTen, personal blog)
- (13) Regarding snorkeling and scuba diving fans, this country is definitely a marine paradise not to *become skipped*. (enTenTen, travel blog)

Both combinations [*become* + participle] are inserted in the modal auxiliary [*be to X*]-construction, which conveys in this case the deontic meaning that 'something should (not) be done by someone'. Since it is stated that the action denoted by the participle has to be performed, this injunction must be addressed to a specific or unspecific agent, which reinforces the activity-reading. In the same vein, a participle such as *hurt* would commonly receive an adjectival interpretation in combination with *be* (without extra context), as in example (14), but when the modal auxiliary [*be to X*]-construction is added in example (15), a processual reading of *hurt* seems more straightforward.

- (14) He is not hurt. ('he is in the state of being unharmed')
- (15) He is not to be hurt. ('you should not hurt him')

In spite of a majority of English participles, Table 7 also contains a few French participles, such as *compté* and *culpabilisé* (examples 16 and 17).

- (16) Tout *devient compté*. Le temps est compté. (frTenTen, 'become counted')
- (17) Peut-on s'interroger sur le profit aux dépens de l'humain sans perpétuellement parler de la solution finale, qui dessert l'interrogation *devenue culpabilisée* parce que des milliers de morts l'emportent sur des exclusions de l'entreprise et du capitalisme ? (frTenTen, book club, 'become made guilty')

Next, the distribution of this measure in our sample is visualized by means of a frequency polygon graph (Figure 5). As for the frequency polygon analysis in Section 4.1, the plot takes into account the token frequencies of the different participle types. The scope of the x-axis is again reduced (from ca. 1110 to 125), in order to focus on the main part of the distribution, excluding 294 outliers.

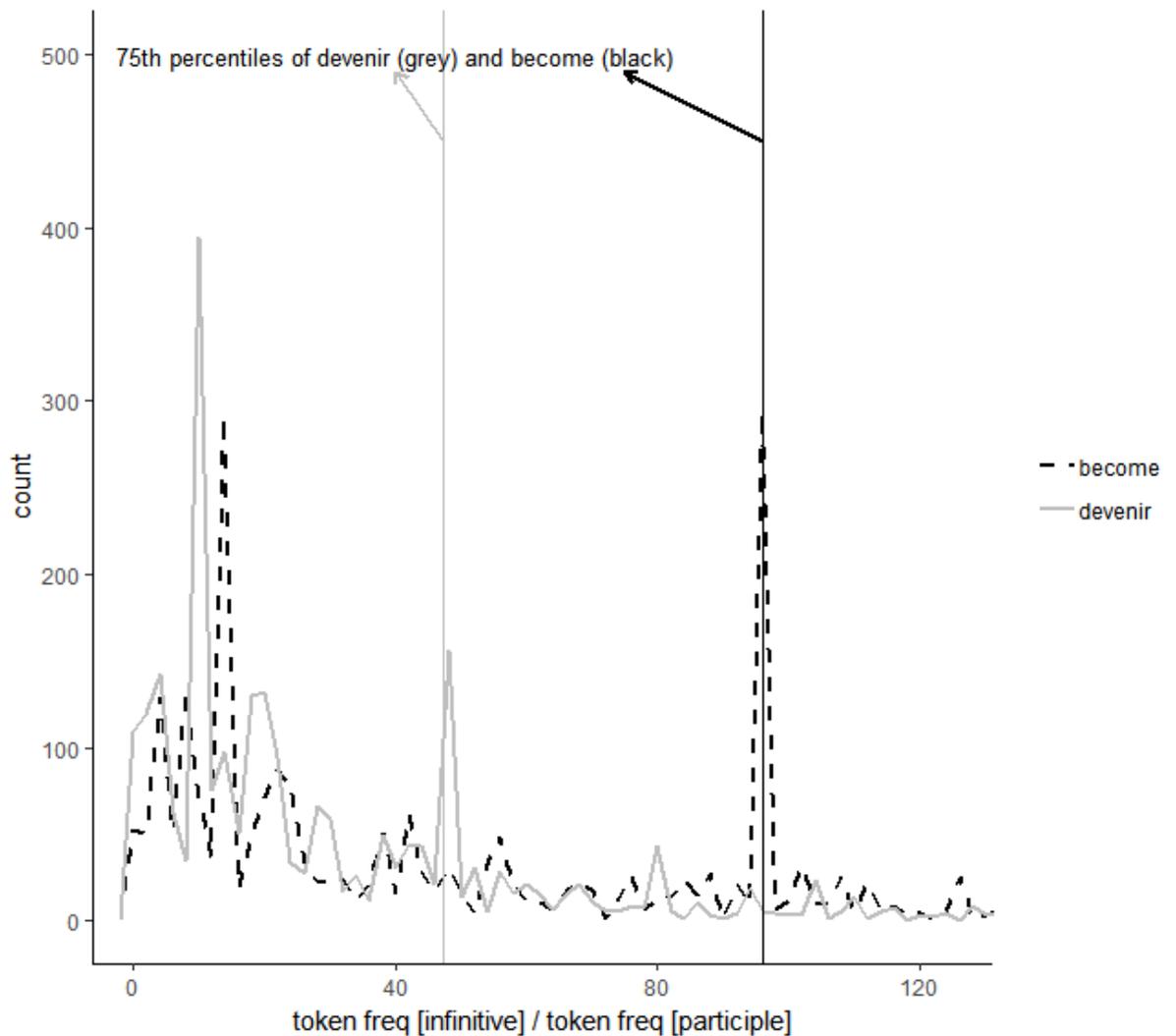


Figure 4. Frequency polygon graph with comparison of *become* and *devenir* for the ratio ‘token freq [infinitive] / token freq [participle]’.

The frequency polygon graph shows that the distributions of both verbs are fairly similar in the lowest interval [0 ; 40], although *devenir* is slightly better represented. However, in the segment of relatively high ratio values (> 40), the frequency polygon graph indicates that both verbs differ: *devenir* reaches its most important peak earlier than *become*, which is also reflected in a higher 75th percentile value for *become* (95.9) than for *devenir* (47.3).

The type *known* is responsible for the major final peak in the distribution of *become* (288 tokens), whereas the major final peak in the distribution of *devenir* is due to its quasi-equivalent in French, the type *connu* (143 tokens). Both participles derive from a cognitive verb (*know / connaître*). As such, it is not the type of participle that is conducive to a passive interpretation in case of a structural ambiguity between the copular and the passive construction. According to this measure, *connu* is thus less verbal than *known*. This is also reflected in the ratio

value that evaluates the compatibility with a degree adverb: *connu* (1.56) scores higher than *known* (0.03 ; ?*very known*). At the same time, *known* is less verbal than a second equivalent in French, namely *su* (cf. Table 7, one token). This makes *known* an interesting intermediate case between the more adjectival *connu* and the more verbal *su*. Hence, *known* could possibly constitute a historical bridging context that paved the way for more verbal participles to combine with *become*.

Table 8 provides the token frequency, the type frequency and the hapax frequency for three intervals of ratio values, namely [0 ; 75th percentile of *devenir* (47.3)], [75th percentile of *devenir* (47.3) ; 75th percentile of *become* (95.9)] and [75th percentile of *become* (95.9) ; ∞].

Table 8. Productivity measures per interval of ratio values and per verb.

verb	interval	token frequency (within sample)	type frequency	hapax frequency
become	[0 ; 47.3]	1432	333	165
devenir	[0 ; 47.3]	1875	439	213
become]47.3 ; 95.9]	698	135	73
devenir]47.3 ; 95.9]	458	146	84
become]95.9 ; ∞]	370	170	113
devenir]95.9 ; ∞]	167	89	55

Become takes the lead in the segment]47.3 ; 95.9] from the perspective of the token frequency, but this tendency is not confirmed by the type frequency and hapax frequency measures, which suggest it is actually *devenir* that slightly outperforms *become* in this intermediate segment of ratio values. The main difference between *become* and *devenir* in this segment is due to a higher token frequency for the most frequent type attested for *become* (*known* with 288 tokens vs *connu* with 143 tokens). Next to more conventionalized combinations such as *devenir permis* (‘become allowed’, 12 tokens) and *devenir pollué* (‘become polluted’, 7 tokens), examples (18) and (19) illustrate that the hapaxes combining with *devenir* in this segment are often rather occasional uses.

- (18) La prise de conscience du retard accumulé vis à vis de nos voisins (de l'OCDE) sur tous les fronts, depuis 26 ans est désormais générale et n'est plus acceptée avec fatalité. Elle *devient refusée*. Donc chacun y travaille. (frTenTen, blog Le Monde, ‘become refused’)
- (19) C'est là aussi que le calcul de la pension est *devenu calculé sur les 6 derniers mois de salaire* comme pour les fonctionnaires, au lieu du dernier salaire. (frTenTen, workers’ party website, ‘become calculated’)

In example (20), the participle *utilisé* (‘used’) is modified by a *by*-phrase expressing the human agent of the action, which underlines the verbal nature of the participle.

- (20) Frappée assez régulièrement tout au long de son développement par une mère despotique qui tient bordel et exploite un hôtel, une patiente me raconte qu'elle *devient utilisée par celle-ci* et dès la puberté pour l'attrait marqué que son corps suscite auprès des clients en mal de sexe. (frTenTen, personal blog, ‘become used by her’)

In the highest segment of ratio values ($[95.9 ; \infty]$), *become* surpasses *devenir* for all three productivity measures. Table 9 gives an overview of the top 20 most frequent participles, measured within our sample, for the subset of ratio values in the interval ($[95.9 ; \infty]$).

Table 9. Top 20 of highest participle token frequencies (within dataset), restricted to the subset of participles with high values for the ratio ‘token freq [infinitive] / token freq [participle]’.

rank	verb	past participle	ratio	token frequency (within sample)
1	become	engaged	102.6	28
2	become	educated	112.1	19
3	become	blocked	125.5	17
4	devenir	distrain (‘distracted’)	109.4	12
5	become	relaxed	143.9	10
6	become	empowered	108.2	10
5	become	absorbed	100.4	10
7	devenir	dépassé (‘exceeded’, ‘outdated’)	103.3	9
8	become	accepted	148.1	8
9	become	possessed	384.4	6
10	become	transformed	134.5	6
11	devenir	mordu (‘bitten’)	128.9	6
12	become	converted	107.8	6
13	become	focused	105.3	6
14	become	covered	103.0	6
15	become	invested	213.5	5
16	become	worried	151.7	5
17	become	incorporated	119.3	5
18	devenir	abusé (‘excessive’)	103.3	5
19	become	broken	96.8	5
20	devenir	possédé (‘possessed’)	371.1	4

In accordance with our initial hypotheses, this top 20 is also dominated by *become*. However, it must be noted that some participles such as *relaxed* and *worried* are derived from psych verbs, which, as already stated above, are in principle less prone to a passive interpretation. This implies that our operationalization is not completely adequate to serve the purpose of detecting the most verbal participles. Clearly, participles conveying psychological states,

very prone to adjectival uses, can still have frequent infinitival counterparts quite independently.

Examples (21) to (23) exemplify these frequent combinations between *become* and a participle with relatively high verbal potential.

- (21) The red wine will *become absorbed* by your enamel and can stain your teeth over time. (enTenTen, resort website)
- (22) Orthodox science is constantly being changed as new theories *become accepted* by the sheeple. (enTenTen, personal blog)
- (23) Then, the masses cling to the big lie as fact, the new gospel for the social order *becomes transformed* into a somber reality. (enTenTen, extreme political essays)

In these instances, the use of *become* seems fairly close to the passive construction. If *become* in the examples above is replaced by *be*, which corresponds to the traditional auxiliary of the passive in English, several elements concur to indicate a passive reading of *be*, rather than a copular interpretation. In addition to the intrinsic verbal nature of the participle, the *by*-phrase and the presence of an indirect complement typical of the verbal use of the participle (cf. *transformed* + PP [*into* X]) trigger in the case of *be* a reading of *be* as passive auxiliary. Insofar the same reasoning can be applied to *become*, these elements also suggest a quasi-passive reading for *become* in the corpus examples cited above. Moreover, the non-resultative aspect of *become* in these examples (cf. present and future tense), focusing on the process rather than the final state, reinforces this quasi-passive interpretation. If *become* is used instead of *be*, this seems to add the notion of change: the process implied by the participle is new and did not happen before. Since these combinations [*become* + participle with high verbal potential] appear to be conventionalized cases, they hint at an ongoing shift towards a structural anchoring of a passive auxiliary-like use of *become*.

5. Conclusion

To conclude, the following two main observations can be retained from this study. First, the subschemas [*become* + past participle] and [*devenir* + past participle] behave quite similarly when it comes to productivity (type frequency, hapax frequency) at the subject complement level, measured for equal sample sizes. Importantly, this does not imply that the subject complements of both verbs are distributed exactly in the same way over the semantic space. However, assuming that the productivity level of the subschema [*become* + past participle] is relatively high, the observation that *become* and *devenir* are almost equally productive

constitutes evidence against the hypothesis, put forward by Guehria (2011), that *devenir* is largely incompatible with past participles.

An inquiry into the adjectival (and, conversely, verbal) potential of the past participles combining with *become* and *devenir* has revealed that (i) *devenir* is better represented in the higher segment of ratios measuring the possibility of modification by a degree adverb and that (ii) *become* is better represented in the higher segment of ratios measuring the frequency of the past participle form in comparison with the inherently verbal infinitival form. While observation (i) indicates a higher adjectival potential for *devenir*, observation (ii) suggests a higher verbal potential for *become*. This also ties neatly in with the hypothesis that, overall, *become* is endowed with a higher degree of passivity, assuming that more eventive participles can trigger more easily a passive reading. However, it must be added that *devenir* is rather type frequent in the intermediate segment of ratio (ii) and is also somewhat represented in the highest segment of ratio (ii). This might indicate an emerging productivity in this area for *devenir* as well.

Finally, in addition to the observations made concerning the overall profile of both verbs, it is noticeable that certain individual examples involving [*become* / *devenir* + past participle] seem very close to a passive construction.

(24)

Further research is needed to clarify (i) to what extent native speakers rate these examples as acceptable and (ii) to what extent they receive a passive reading. Other research avenues can be explored as well, such as a synchronic inquiry into how the subschema [*become* / *devenir* + past participle] relates to recognized (quasi-)passive constructions (e.g. *get*-passive) or a diachronic analysis that examines to which degree this compatibility of *become* and *devenir* with past participles is already attested in earlier language stages.

References

- Baayen, R. H. and Lieber, R. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29 (5): 801–844.
- Baayen, R. H. 1993. On frequency, transparency and productivity. In *Yearbook of Morphology 1992*, edited by Geert Booij and Jaap van Marle. Dordrecht: Springer. 181-208.
- Baayen, R. H. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In *Corpus Linguistics*, edited by Anke Lüdeling, Merja Kytö, and Merja Kytö. Berlin: de Gruyter. 899–919.
- Baroni, M. and Evert, S. 2014. The ZipfR Package for Lexical Statistics: A Tutorial Introduction. Available at <http://zipfr.r-forge.r-project.org/materials/zipfrtutorial.pdf> [last accessed 1 June 2019].

- Craig, C. G. 1991. Ways to go in Rama: A case study in polygrammaticalization. In *Approaches to Grammaticalization. Volume II. Types of grammatical markers*, edited by Elizabeth Closs Traugott and Bernd Heine. Amsterdam: John Benjamins Publishing Company. 455-492.
- De Sutter, G. 2005. Rood, Groen, Corpus! Een Taalgebruiksgebaseerde Analyse van Woordvolgordevariatie in Tweeledige Werkwoordelijke Eindgroepen. PhD Thesis, Katholieke Universiteit Leuven.
- Dik, S. C. 1987. Copula Auxiliariation: How and Why? In *Historical Development of Auxiliaries*, edited by Martin Harris and Paolo Ramat. Berlin: de Gruyter. 53–84.
- Gaeta, L. and Ricca, D. 2006. Productivity in Italian Word Formation: A Variable-Corpus Approach. *Linguistics* 44 (1): 57–89.
- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Guehria, W. 2011. La Structure Attributive avec *devenir* comme Construction Marquée dans l'Ensemble Sous-déterminé des Phrases de Forme N V_{état} Adj. *Langue française* 171 (3): 135–146.
- Hopper, P. J. and Thompson, S. A. 1980. Transitivity in Grammar and Discourse. *Language* 56 (2): 251–299.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. 2014. The Sketch Engine: Ten Years On. *Lexicography* 1 (1): 7–36.
- Laca, B. 2000. Auxiliariation et Copularisation. *Revue de Linguistique Romane* 64: 427–443.
- Raineri, S. 2010. Analyse Contrastive Français-Anglais du Passif dans une Perspective Constructionnelle : Sens et Fonction de *BE Ven*, *ETRE Vé*, *GET Ven* et *SE FAIRE Ver*. PhD Thesis, Université de la Sorbonne nouvelle - Paris III.
- Sansò, A. and Ramat, A. G. 2015. Deictic Motion Verbs as Passive Auxiliaries: The Case of Italian *andare* 'go' and *venire* 'come'. *Transactions of the Philological Society* 114 (1): 1–24.
- Sleeman, P. 2014. From Participle to Adjective in Germanic and Romance. In *Adjectives in Germanic and Romance*, edited by Petra Sleeman, Freek Van de Velde and Harry Perridon. Amsterdam: John Benjamins Publishing Company. 171–198.
- Wickham, H. 2009. *Ggplot2: Elegant graphics for data analysis*. New York: Springer.
- Zeldes, A. 2012. *Productivity in Argument Selection from Morphology to Syntax*. Berlin: De Gruyter.